

WHAT IS CLAIMED IS:

1. A method of processing a document page comprising the steps of:

a) receiving image data representing at least a portion of a first side of said document page;

b) extracting a set of connected components from said first side image data;

d) erasing pixels representing said set of connected components from said first side image data; and

e) extracting a first set of features from said first side image data as processed by said d) step, whereby said first set of features represent a bleed-through image of at least a portion of a second side of said document page.

2. The method of claim 1 wherein said step of extracting said first set of features comprises flipping said first side image data as processed by said d) step.

3. The method of claim 1 further comprising the steps of:

f) scanning said second side of said document page to capture second side image data;

g) removing pixels representing said set of connected components from said second side image data; and

h) extracting a second set of features from said second side image data as processed by said g) step.

4. The method of claim 3 further comprising the step of:

i) verifying said second side of said document page as a match to said first side of said document page based on a match between said first set of features and said second set of features.

5. The method of claim 4 wherein a Hausdorff distance metric is used to establish a match between said first set of features and said second set of features.

7. The method of claim 4 wherein a Euclidean distance metric is used to establish a match between said first set of features and said second set of features.

1 8. The method of claim 1 wherein said extracting step comprises  
2 extracting CCITT Group IV pass codes.

1 9. The method of claim 1 wherein said extracting step comprises  
2 extracting word length sequences.

1 10. The method of claim 1 wherein said receiving step comprises  
2 scanning in said first side of said document page.

1 11. A method for processing a multi-page document comprising the  
2 steps of:

3 a) scanning in front sides of a plurality of pages of said multi-page  
4 document to develop image data for each of said front sides;

5 b) for each of said front sides, identifying in said image data, regions  
6 corresponding to printed matter on said front sides;

7 c) developing a bleed-through image for each of said front sides;

8 d) scanning in back sides of said plurality of pages of said multi-page  
9 document to develop image data for each of said back sides; and

10 e) matching said image data for said back sides to said bleed-through  
11 image data to identify matching front and back sides.

1 12. The method of claim 11 wherein said bleed-through image includes  
2 printed regions not occluded by said regions corresponding to printed matter of said front  
3 sides

1 13. The method of claim 11 further comprising the step of:

2 f) controlling reproduction of said multi-page document responsive to  
3 results of said matching step.

1 14. The method of claim 11 further comprising the step of:

2 f) controlling faxing of said multi-page document responsive to results  
3 of said matching step.

1 15. A document processing system comprising:  
2 a scanner that scans in front sides of a plurality of pages of said multi-page  
3 document to develop image data for each of said front side and scans in back sides of  
4 said plurality of pages of said multi-page document to develop image data for each of  
5 said back sides;  
6 an image processing system that  
7 a) for each of said front sides, identifies in said image  
8 data, regions corresponding to printed matter on said front sides;  
9 b) develops a bleed-through image for each of said front  
10 sides; and  
11 c) matches said image data for said back sides to said  
12 bleed-through image data to identify matching front and back sides.

1 16. The document processing system of claim 15 wherein said bleed-  
2 through image includes printed regions not occluded by said regions corresponding to  
3 printed matter of said front sides.

1 17. The document processing system of claim 15 wherein said image  
2 processing system further controls document reproduction responsive to said matching  
3 front and back sides.

1 18. The document processing system of claim 15 wherein said image  
2 processing system further controls document faxing responsive to said matching front and  
3 back sides.

1 19. A method for copying a multi-page document comprising the steps  
2 of:  
3 scanning in a plurality of document pages to capture image data for each  
4 of said plurality of pages;  
5 storing said image data for each of said plurality of pages; and  
6 comparing said image data for each of said plurality of pages to identify  
7 twice scanned pages.

a

add B27